ELSEVIER
Theoretical Computer Science 156 (1996) 145–158
Theoretical Computer Science

Stochastically costed tree automata: Turakainen's theorem

Olympia Louscou-Bozapalidou
Posidonos 7, Nea Krini, 55132 Thessaloniki, Greece

Received June 1994; revised December 1994
Communicated by A. Salomaa

Abstract

We establish Turakainen's theorem in the framework of trees; precisely, we show that every forest of the form $\{t \mid C_{\mathscr{A}}(t) > 0\}$ ($\mathscr{A}$ an $\mathbb{R}$-costed tree automaton) coincides with a stochastic forest, i.e. a forest of the form $\{t \mid C_{\mathscr{P}}(t) > \lambda\}$ with $\mathscr{P}$ an actual double stochastic tree automaton and λ a cut-point.

As an application we obtain some remarkable closure properties concerning stochastic forests and languages.

0. Introduction

Turakainen has shown that the "positive part" of a rational series r, i.e. the language $\{w \mid r(w) > 0\}$, coincides modulo the empty word with a language $\{w \mid s(w) > \lambda\}$ where s is the function computed by an actual double stochastic word automaton and λ is a suitable cut-point (cf. [8, 10]).

This result is an effective tool used to get some interesting closure properties for the hardly manipulating family of stochastic languages.

The purpose of this paper is to transfer Turakainen's theorem in a higher complexity level, that of trees.

Stochastic tree automata were introduced in [6] where the basic properties of the functions they compute are examined; it is shown, for instance, that stochastic tree functions form a convex subset of the space of all tree functions as well as they are closed with respect to composition with a tree homomorphism.

Also, a stochastic forest, i.e. a forest of the form $\{t \mid f(t) > \lambda\}$ with f stochastic tree function, collapses to a recognizable one, whenever its cut-point λ is isolated [6].

The present paper is divided into three sections. In Section 1, we prove two (necessary in the sequel) results concerning costed tree automata (also called linear representations in [1]).

The first one states that the function $G: T_\Sigma \to \mathbb{R}$ computed by such a tree automaton is upper bounded by a (tree) geometric progression, that is the following

SSDI 0304-3975(95)00009-7

holds:

$$|G(t)| < M^{1+\mathrm{size}(t)}, \quad M \text{ real constant.}$$

As for the second result, it ensures that the support of a **Q**-costed tree function equals the support of a **Z**-costed tree function.

In Section 2 the next important result is established: from any costed tree automation $\mathscr{A}$ we can construct an actual double stochastic tree automaton $\mathscr{P}$ such that

$$C_{\mathscr{A}}(t) = m^{\|t\|}[C_{\mathscr{P}}(t) - \lambda] \quad \text{for all } t \in T_\Sigma - \Sigma_0,$$

where m and $0 \leqslant \lambda \leqslant 1$ are appropriately chosen constants and $\|t\|$ denotes the number of symbols of $\Sigma - \Sigma_0$ occurring in t. We immediately deduce that the three families of forests

- stochastic
- double stochastic and
- positive parts of costable tree functions

are identical.

Then we prove (Section 3) that $\mathbb{K}$-STOCH, the family of $\mathbb{K}$-stochastic forests, is closed under top-catenation and intersection with supports of $\mathbb{K}$-costed tree functions, as well as under union with **Q**-supports.

Applying the frontier operator, from the above results we yield interesting "byproducts" such as: the catenation of a $\mathbb{K}$-stochastic language, with $\mathbb{K}$-rational langauge, is a frontier of a $\mathbb{K}$-stochastic forest.

As usual, T_Σ denotes the set of trees over the ranked alphabet $\Sigma = \Sigma_0 \cup \Sigma_1 \cup \cdots \cup \Sigma_N$; subsets of T_Σ are termed forests (over Σ). For details we refer to [3, 5].

1. Tree automata with costs

Given a finite ranked alphabet Σ and a ring $\mathbb{K}$, a Σ-tree automaton with costs in $\mathbb{K}$, is a triple

$$\mathscr{A} = (Q, \alpha, f)$$

formed by a finite set Q (the states) and a function $f: Q \to \mathbb{K}$ (final costs); the moves of $\mathscr{A}$ are described by the family of functions

$$\alpha_\sigma : Q^n \to \mathbb{K}^Q \quad (\sigma \in \Sigma_n,\ n \geqslant 0).$$

For $q_1, \ldots, q_n$, $q \in Q$ and $\sigma \in \Sigma_n$, the scalar $\alpha_\sigma(q_1, \ldots, q_n)(q)$ is the cost of the move $q_1 \ldots q_n \xrightarrow{\sigma} q$. The α_σ's are naturally extended into functions

$$\bar{\alpha}_\sigma : (\mathbb{K}^Q)^n \to \mathbb{K}^Q \quad (\sigma \in \Sigma_n,\ n \geqslant 0).$$

by setting

$$\bar{\alpha}_\sigma(x_1, \ldots, x_n)(q) = \sum_{q_1, \ldots, q_n \in Q} x_1(q_1) \ldots x_n(q_n)\alpha_\sigma(q_1, \ldots, q_n)(q), \quad x_i \in \mathbb{K}^Q \ (1 \leqslant i \leqslant n).$$

The reachability map $h_{\mathscr{A}}: T_\Sigma \to \mathbb{K}^Q$ of $\mathscr{A}$ is given by the inductive formula

$$h_{\mathscr{A}}(\sigma(t_1, \ldots, t_n)) = \bar{\alpha}_\sigma(h_{\mathscr{A}}(t_1), \ldots, h_{\mathscr{A}}(t_n))$$

where $\sigma \in \Sigma_n$ $(n \geqslant 0)$ and $t_j \in T_\Sigma$.

The tree function computed by $\mathscr{A}$, $C_{\mathscr{A}}: T_\Sigma \to \mathbb{K}$ is then

$$C_{\mathscr{A}}(t) = \sum_{q \in Q} h_{\mathscr{A}}(t)(q) \cdot f(q).$$

Two such automata $\mathscr{A}$ and $\mathscr{B}$ are said to be equivalent, if $C_{\mathscr{A}} = C_{\mathscr{B}}$.

Call a function $G: T_\Sigma \to \mathbb{K}$ costable if $G = C_{\mathscr{A}}$, for some $\mathbb{K}$-costed Σ-tree automaton $\mathscr{A}$. When taking costs in the field of real numbers $\mathbb{R}$, the absolute value of the values of a costable function is exponentially upper bounded with respect to tree size; precisely:

Proposition 1. *For each costable function $G: T_\Sigma \to \mathbb{R}$ we can find a real number $M > 0$ such that*

$$|G(t)| < M^{1+\mathrm{size}(t)} \quad \textit{for all } t \in T_\Sigma,$$

where size(t) is the number of symbols of Σ occurring in t.

Proof. Consider an $\mathbb{R}$-costed Σ-tree automaton $\mathscr{A} = (Q, \alpha, f)$ such that $G = C_{\mathscr{A}}$ and put

$$N = \max\{|\alpha_\sigma(q_1, \ldots, q_n)(q)|, |f(q)| \mid \sigma \in \Sigma_n \ (n \geqslant 0) \text{ and } q_1, \ldots, q_n, q \in Q\}$$

and

$$\kappa = \mathrm{card}\, Q^{\deg \Sigma},$$

where $\deg \Sigma$ denotes the maximal index μ such that $\Sigma_\mu \neq \emptyset$ (called the degree of the alphabet Σ). It holds that

$$|h_{\mathscr{A}}(t)(q)| < (\kappa N)^{\mathrm{size}(t)} \quad \text{for all } t \in T_\Sigma.$$

The desired inequality follows by taking $M = \kappa N$. $\square$

The result below has to do with $\mathbf{Q}$-costed functions ($\mathbf{Q}$ is the field of rational numbers).

Proposition 2. *For any costable function $G: T_\Sigma \to \mathbf{Q}$, there exists a positive integer m such that the function*

$$t \mapsto G(t) \cdot m^{1+\mathrm{size}(t)}$$

takes integer values and is costable.

Proof. Assume that $G = C_{\mathscr{A}}$ for a $\mathbf{Q}$-costed Σ-tree automaton $\mathscr{A} = (Q, \alpha, f)$ and let m be the least common multiple of the denominators of the next finite list of rational numbers

$$\alpha_\sigma(q_1,\ldots,q_n)(q), f(q) \quad \text{with } \sigma\in\Sigma_n\ (n\geqslant 0),\ q_i, q\in Q.$$

Then the $\mathbf{Z}$-costed Σ-tree automaton $\mathscr{B}=(Q,\beta,g)$ with

$$\left.\begin{array}{l}\beta_\sigma(q_1,\ldots,q_n)(q)=m\cdot\alpha_\sigma(q_1,\ldots,q_n)(q),\\ g(q)=m\cdot f(q),\end{array}\right\}\quad \sigma\in\Sigma_n\ (n\geqslant 0)\ \ q_i, q\in Q.$$

has the announced property. □

Corollary. *For any costable function $G: T_\Sigma\to\mathbf{Q}$, we can determine a positive rational δ with the property: for any tree t lying on the support of G*

$$|G(t)|>\delta^{1+\mathrm{size}(t)}$$

recall that $\mathrm{supp}(G)=\{t\mid G(t)\neq 0\}$.

Remark. Costable functions were introduced in [1] by means of an equivalent formalism; the reader is referred to this work for additional information.

2. Stochastically costed tree automata

We fix our finite ranked alphabet Σ.

A stochastic tree automaton (STA) is an $\mathbb{R}$-costed Σ-tree automaton $\mathscr{P}=(Q,\pi,\eta)$, where the function $\eta: Q\to\mathbb{R}$ takes exclusively the values 0 or 1 and the move functions

$$\pi_\sigma: Q^n\to\mathbb{R}^Q$$

are subjected to verify the conditions below:

(s_1) for all $q_1,\ldots,q_n$, $q\in Q$ and $\sigma\in\Sigma_n$ $(n\geqslant 0)$,

$$\pi_\sigma(q_1,\ldots,q_n)(q)\geqslant 0,$$

(s_2) for all $q_1,\ldots,q_n\in Q$ and $\sigma\in\Sigma_n$ $(n\geqslant 0)$,

$$\sum_{q\in Q}\pi_\sigma(q_1,\ldots,q_n)(q)=1.$$

Proposition 3 (cf. Louscou-Bozapalidou [6]). *For any tree $t\in T_\Sigma$, $\pi_{\mathscr{P}}(t)$ is a stochastic Q-vector and $0\leqslant c_{\mathscr{P}}(t)\leqslant 1$, where $\pi: T_\Sigma\to\mathbb{R}^Q$ is the reachability map of $\mathscr{P}$ and $C_{\mathscr{P}}$ the function computed by $\mathscr{P}$.*

Forests of the form

$$F(\mathscr{P},\lambda)=\{t\in T_\Sigma\mid C_{\mathscr{P}}(t)>\lambda,\ 0\leqslant\lambda\leqslant 1\}$$

are termed stochastic.

The family of such forests properly contains that of recognizable forests, whereas

Theorem 1 (Louscou-Bozapalidou [6]). *If λ is an isolated cut point of the STA $\mathscr{P}$ (i.e. there exists $\varepsilon > 0$ such that $|C_{\mathscr{P}}(t) - \lambda| \geqslant \varepsilon$ for all $t \in T_\Sigma$), then $F(\mathscr{P}, \lambda)$ is a recognizable forest.*

A stochastic tree automaton $\mathscr{P} = (Q, \pi, \eta)$ is said to be

(i) actual, if all numbers in (s_1) are strictly positive,

(ii) double stochastic, if next condition is fulfilled:

(s_3) $\sum_{p \in Q} \pi_\sigma(q_1, \ldots, q_{i-1}, p, q_{i+1}, \ldots, q_n)(q) = 1$

for any choice $q_j, q \in Q$ $(j \neq i)$ and any place i $(1 \leqslant i \leqslant n)$.

Remarks. (1) We should notice that stochastic tree automata are the natural generalization of string stochastic automata (cf. [8]); to see this we only have to consider monadic ranked alphabets with just one symbol of rank 0.

(2) The probabilistic sinking automata of Magitor and Moran (cf. [7]) differ from the ours on axioms (s_2) and (s_3); the corresponding axiom used there is the "global summability" condition

$$\sum_{q_1, \ldots, q_n \in Q} \pi_\sigma(q_1, \ldots, q_n)(q) = 1.$$

(3) STAs can be defined relative to any subfield $\mathbb{K}$ of the reals $\mathbb{R}$; we then speak of $\mathbb{K}$-STAs, etc.

Our objective in the present paper is to transfer an important theorem due to Turakainen (cf. [8, 10]) into the framework of trees.

Theorem 2. *Let $\mathbb{K}$ be a subfield of $\mathbb{R}$. From each $\mathbb{K}$-costed Σ-tree automaton $\mathscr{A} = (Q, \alpha, f)$ we can construct an actual double stochastic Σ-tree automaton $\mathscr{P} = (Q', \pi, n)$ with* $\operatorname{card} Q + \operatorname{card} \Sigma_0 + 3$ *states, such that for all trees $t \in T_\Sigma - \Sigma_0$,*

$$C_{\mathscr{A}}(t) = m^{\|t\|}(C_{\mathscr{P}}(t) - \lambda),$$

where m and $0 \leqslant \lambda \leqslant 1$ are effectively determined constants and $\|t\|$ denotes the number of symbols of $\Sigma - \Sigma_0$ occurring in $t \in T_\Sigma$.

We need two auxiliary results.

Lemma 1. *Any $\mathbb{K}$-costed Σ-tree automaton $\mathscr{A} = (Q, \alpha, f)$ is equivalent to another one $\mathscr{B} = (\bar{Q}, \beta, \bar{f})$ satisfying the conditions*

$$\sum_{r \in Q} \beta_\sigma(\ldots, r_{i-1}, r, r_{i+1}, \ldots)(r') = 0 \quad \textit{at any place } i \ (1 \leqslant i \leqslant n)$$

and

$$\sum_{r \in Q} \beta_\sigma(r_1, \ldots, r_n)(r) = 0$$

for all $\sigma \in \Sigma_n$ $(n \geqslant 1)$ *and* $r', r, r_j \in \bar{Q}$.

Proof. We adjoin to Q two new states p_0, p_+:

$$\bar{Q} = Q \cup \{p_0, p_+\}$$

and define for all $\sigma \in \Sigma_n$ $(n \geqslant 1)$:

$$\beta_\sigma(q_1, \ldots, q_n)(q) = \alpha_\sigma(q_1, \ldots, q_n)(q) \quad \text{if all } q_i, q \in Q,$$

$$\beta_\sigma(q_1, \ldots, q_n)(p_+) = -\sum_{q \in Q} \alpha_\sigma(q_1, \ldots, q_n)(q) \quad \text{if all } q_i \in Q,$$

$$\beta_\sigma(\overbrace{\ldots, \underset{\substack{\uparrow \\ i_1\text{th place}}}{p_0}, \ldots, \underset{\substack{\uparrow \\ i_\kappa\text{th place}}}{p_0}, \ldots}^{\text{elements of } Q})(p) = \begin{cases} (-1)^\kappa \sum\limits_{q_{i_l} \in Q} \alpha_\sigma(\ldots, q_{i_1}, \ldots, q_{i_\kappa}, \ldots)(p) & \text{if } p \in Q, \\ (-1)^{\kappa+1} \sum\limits_{q, q_{i_l} \in Q} \alpha_\sigma(\ldots, q_{i_1}, \ldots, q_{i_\kappa}, \ldots)(q) & \text{if } p = p_+. \end{cases}$$

In all remainder cases $\beta_\sigma(p_1, \ldots, p_n) = 0$.

By construction $\mathscr{B}$ satisfies the announced conditions. Further, for each $c \in \Sigma_0$ we set

$$\beta_c(q) = \alpha_c(q) \text{ and } \beta_c(p_0) = \beta_c(p_+) = 0, \qquad q \in Q$$

and

$$\bar{f}(p_0) = \bar{f}(p_+) = 0, \qquad \bar{f}(q) = f(q), \qquad q \in Q.$$

Using induction on the size of $t \in T_\Sigma$ we show that

$$h_{\mathscr{B}}(t)(q) = h_{\mathscr{A}}(t)(q) \quad \text{for all } q \in Q.$$

In fact, this formula is true for $t = c \in \Sigma_0$, whereas for $t = \sigma(t_1, \ldots, t_n)$ $(\sigma \in \Sigma_n, n \geqslant 1, t_j \in T_\Sigma)$ we have

$$\begin{aligned} h_{\mathscr{B}}(\sigma(t_1, \ldots, t_n))(q) &= \sum_{r_i \in \bar{Q}} h_{\mathscr{B}}(t_1)(r_1) \cdots h_{\mathscr{B}}(t_n)(r_n) \cdot \beta_\sigma(r_1, \ldots, r_n)(q) \\ &= h_{\mathscr{A}}(\sigma(t_1, \ldots, t_n))(q) + A + B, \end{aligned}$$

where A (resp. B) is a sum of terms having a factor of the form $h_{\mathscr{B}}(t_i)(p_0)$ (resp. $\beta_\sigma(\ldots, p_+, \ldots)$) which obviously is equal to zero. Consequently,

$$C_{\mathscr{B}}(t) = \sum_{r \in \bar{Q}} h_{\mathscr{B}}(t)(r) \cdot \bar{f}(r) = \sum_{q \in Q} h_{\mathscr{A}}(t)(q) \cdot f(q) = C_{\mathscr{A}}(t),$$

that is, $C_{\mathscr{B}} = C_{\mathscr{A}}$ and this completes the proof. $\square$

Lemma 2. *Let $\mathscr{A} = (Q, \alpha, f)$ be a $\mathbb{K}$-costed Σ-tree automaton satisfying the following conditions:*

(i) *for all $c \in \Sigma_0$,*

$$\sum_{q \in Q} \alpha_c(q) = 1 \text{ and}$$

(ii) *for all $\sigma \in \Sigma_n$ $(n \geqslant 1)$ and $q_j, p, q \in Q$,*

(1) $\sum_{q \in Q} \alpha_\sigma(\ldots, q_{i-1}, q, q_{i+1}, \ldots)(p) = 0$ *at any place i $(1 \leqslant i \leqslant n)$,*

(2) $\sum_{p \in Q} \alpha_\sigma(q_1, \ldots, q_n)(p) = 0$.

Further given a real number ξ we define the tree automaton $\mathscr{B} = (Q, \beta, f)$ by putting

$$\beta_c = \alpha_c \quad \text{for all } c \in \Sigma_0,$$

$$\beta_\sigma(q_1, \ldots, q_n)(q) = \alpha_\sigma(q_1, \ldots, q_n)(q) + \xi \ (\sigma \in \Sigma_n,\ n \geqslant 1,\ q_i \in Q).$$

Then for all trees $t \in T_\Sigma - \Sigma_0$, the following holds:

$$h_{\mathscr{B}}(t)(q) = h_{\mathscr{A}}(t)(q) + \xi^{\|t\|}(\text{card } Q)^{\|t\| - 1}.$$

Proof. By induction on $\|t\|$; assume that $\|t\| = 1$, this means that

$$t = \sigma(c_1 \ldots c_n) \quad (\sigma \in \Sigma_n,\ c_i \in \Sigma_0)$$

and so

$$\begin{aligned} h_{\mathscr{B}}(t)(q) &= h_{\mathscr{A}}(t)(q) + \left(\sum_{q_1 \in Q} \alpha_{c_1}(q_1)\right) \cdots \left(\sum_{q_n \in Q} \alpha_{c_n}(q_n)\right) \xi \\ &= h_{\mathscr{A}}(t)(q) + \xi, \end{aligned}$$

where assumption (i) is used to deduce the last equality.

Without any loss of generality, we can write any tree $t \in T_\Sigma - \Sigma_0$ in the form

$$t = \sigma(c_1, \ldots, c_\kappa, t_{\kappa+1}, \ldots, t_n)$$

with $\sigma \in \Sigma_n$ $(n \geqslant 1)$ $c_j \in \Sigma_0$ and $0 < \|t_i\| < \|t\|$ for $i = \kappa + 1, \ldots, n$.
Then

$$\begin{aligned} & h_{\mathscr{B}}(t)(q) \\ &= \sum_{q_1, \ldots, q_n \in Q} \alpha_{c_1}(q_1) \cdots \alpha_{c_\kappa}(q_\kappa) [h_{\mathscr{A}}(t_{\kappa+1})(q_{\kappa+1}) + \xi^{\|t_\kappa\|}(\text{card } Q)^{\|t_\kappa\| - 1}] \\ &\quad \cdots [h_{\mathscr{A}}(t_n)(q_n) + \xi^{\|t_n\|}(\text{card } Q)^{\|t_n\| - 1}][\alpha_\sigma(q_1, \ldots, q_n)(q) + \xi] \\ &= h_{\mathscr{A}}(t)(q) + \xi^{\|t_{\kappa+1}\| + \cdots + \|t_n\| + 1} (\text{card } Q)^{\|t_{\kappa+1}\| + \cdots + \|t_n\| + (n - \kappa)} \\ &\quad \times (\text{card } Q)^{n - \kappa} + A + B. \end{aligned}$$

Every summand in A has a factor of the form

$$\sum_{q \in Q} h_{\mathscr{A}}(t_i)(q)$$

which by assumption (ii) (2) equals to zero and every summand in B has a factor of the form

$$\sum_{q \in \bar{Q}} \alpha_\sigma(\ldots, q, \ldots)$$

which again is 0, because of (ii)(1).

Hence, the final sum is equal to

$$h_{\mathscr{A}}(t)(q) + \xi^{\|t\|}(\text{card } Q)^{\|t\|-1}$$

as desired. $\square$

Now, we are ready to give the proof of Theorem 2; it will be done in four steps.

Proof of Theorem 2.
Step 1: We introduce $(\text{card } \Sigma_0) + 1$ new states

$$q_+, q_c \mid c \in \Sigma_0$$

and we define $\mathscr{A}_1 = (Q_1, \alpha_1, f_1)$ by setting

$$Q_1 = Q \cup \{q_+\} \cup \{q_c \mid c \in \Sigma_0\},$$

$$f_1(q_+) = 1, \quad f_1(q) = 0 \qquad \text{for all } q \in Q,$$

$$f_1(q_c) = \begin{cases} 1 & \text{whenever } C_{\mathscr{A}}(c) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover $(\alpha_1)_c = q_c$, for all $c \in \Sigma_0$, while the moves of $\mathscr{A}_1$ are given by

$$(\alpha_1)_\sigma(q_1, \ldots, q_n)(q) = \alpha_\sigma(q_1, \ldots, q_n)(q) \quad \text{if } q, q_i \in Q,$$

$$(\alpha_1)_\sigma(\ldots, q, \ldots, q_c, \ldots)(q') = \bar{\alpha}_\sigma(\ldots, q, \ldots, \alpha_c, \ldots)(q') \quad \text{if } q, q' \in Q,$$

$$(\alpha_1)_\sigma(\ldots, p, \ldots, q_{c'}, \ldots)(q_+) = \sum_{q \in Q} \bar{\alpha}_\sigma(\ldots, p, \ldots, \alpha_c, \ldots)(q) \cdot f(q).$$

In all other cases $(\alpha_1)_\sigma$ equals 0.

A straightforward calculation shows that

$$C_{\mathscr{A}_1}(t) = \begin{cases} 0 & \text{if } t = c \in \Sigma_0 \text{ and } C_{\mathscr{A}}(c) = 0, \\ 1 & \text{if } t = c \in \Sigma_0 \text{ and } C_{\mathscr{A}}(c) > 0, \\ C_{\mathscr{A}}(t) & \text{if } t \in T_\Sigma - \Sigma_0. \end{cases}$$

Step 2: Applying Lemma 1 we get from $\mathscr{A}_1$ an equivalent ($\mathbb{K}$-costed) tree automaton $\mathscr{A}_2 = (Q_2, \alpha_2, f_2)$ with $Q_2 = Q_1 \cup \{p_0, p_+\}$:

$$f_2(p_0) = f_2(p_+) = 0 \text{ and } f_2(p) = f_1(p) \qquad \text{for all } p \in Q_1,$$

$$(\alpha_2)_c = q_c \quad \text{for all } c \in \Sigma_0$$

and α_2 verifies the conditions stated in that Lemma.

Step 3: Since $\mathscr{A}_2$ satisfies the assumptions of Lemma 2, we obtain a new ($\mathbb{K}$-costed) tree automaton $\mathscr{A}_3 = (Q_3, \alpha_3, f_3)$ with

$$Q_3 = Q_2, \quad f_3 = f_2, \quad (\alpha_3)_c = q_c \text{ for all } c \in \Sigma_0$$

and

$$(\alpha_3)_\sigma(p_1, \dots, p_n)(p) = (\alpha_2)_\sigma(p_1, \dots, p_n)(p) + \xi,$$

where ξ is chosen to be so large that all costs $(\alpha_3)_\sigma(p_1, \dots, p_n)(p)$ are positive. Then for all $t \in T_\Sigma - \Sigma_0$ the following holds:

$$h_{\mathscr{A}_3}(t)(p) = h_{\mathscr{A}_2}(t)(p) + (card\, Q_3)^{\|t\|-1}\, \xi^{\|t\|}$$

Step 4: Finally, let $\mathscr{A}_4 = (Q_4, \alpha_4, f_4)$ be the ($\mathbb{K}$-costed) tree automaton having the same state set, final set and constants as $\mathscr{A}_3$:

$$Q_4 = Q_3, \quad f_4 = \{q_+\} \cup \{q_c \mid C_{\mathscr{A}}(c) > 0\}, \quad (\alpha_4)_c = (\alpha_3)_c, \; c \in \Sigma_0.$$

Its moves are defined by

$$(\alpha_4)_\sigma(p_1, \dots, p_n)(p) = \frac{1}{(\text{card}\, Q_4)\,\xi}(\alpha_3)_\sigma(p_1, \dots, p_n)(p)$$

Then $\mathscr{A}_4$ is by construction an actual double stochastic tree automaton $\mathscr{P}$ and for $t \in T_\Sigma - \Sigma_0$,

$$h_{\mathscr{P}}(t)(p) = \frac{1}{(\text{card}\, Q_4)^{\|t\|}\, \xi^{\|t}}\, h_{\mathscr{A}_3}(t)(p).$$

Taking into account that $\text{card}\, Q_4 = \text{card}\, Q + \text{card}\, \Sigma_0 + 3$, and

$$\sum_{p \in Q_4} f_4(p) = \gamma \leqslant (\text{card}\, \Sigma_0) + 1,$$

we get that for all $t \in T_\Sigma - \Sigma_0$,

$$C_{\mathscr{P}}(t) = \frac{1}{[\xi\, \text{card}\, Q_4]^{\|t\|}}\, C_{\mathscr{A}}(t) + \frac{1}{\text{card}\, Q_4}$$

or

$$C_{\mathscr{P}}(t) = \frac{1}{m^{\|t\|}}\, C_{\mathscr{A}}(t) + \lambda$$

with $m = \xi\, \text{card}\, Q_4$ and $\lambda = \gamma/\text{card}\, Q_4$. $\square$

The positive part of a costable function $G: T_\Sigma \to \mathbb{R}$ is defined to be the forest

$$\{t \in T_\Sigma \mid G(t) > 0\}.$$

Corollary. *The three families below*

(i) *$\mathbb{K}$-stochastic forests,*

(ii) *double $\mathbb{K}$-stochastic forests and*

(iii) *positive parts of $\mathbb{K}$-costable functions mutually coincide and we denote by $\mathbb{K}$-STOCH this common family.*

Proof. One direction is easy: if

$$F = \{t \in T_\Sigma \mid C_{\mathscr{P}}(t) > \lambda\}$$

is a stochastic forest, then

$$F = \{t \in T_\Sigma \mid C_{\mathscr{A}}(t) > 0\}$$

with $\mathscr{A} = \mathscr{P} - \lambda$. To prove the converse, we consider an arbitrary $\mathbb{R}$-costed tree automaton $\mathscr{A} = (Q, \alpha, f)$ and the associated actual double stochastic $\mathscr{P}$ having the property of Theorem 2. Then for $t \in T_\Sigma - \Sigma_0$,

$$C_{\mathscr{P}}(t) > \lambda \text{ iff } C_{\mathscr{A}}(t) > 0.$$

On the other hand, if $c \in \Sigma_0$ is such that $C_{\mathscr{A}}(c) > 0$, then $C_{\mathscr{P}}(c) = 1 > \lambda$ and vice versa; in other words,

$$F(\mathscr{P}, \lambda) = \{t \mid C_{\mathscr{A}}(t) > 0\}. \qquad \square$$

3. Applications

In this section we apply our Theorem 2 (or rather its corollary) in order to get some closure properties of stochastic forests; then, projecting by the frontier operator we obtain nice results concerning languages.

By $\mathbb{K}$-SUPP we denote the family of supports of costable functions $G: T_\Sigma \to \mathbb{K}$; we simply write SUPP for the case $\mathbb{K} = \mathbb{R}$.

Proposition 4. *It holds that*

$$\mathbb{K}\text{-}SUPP \subseteq \mathbb{K}\text{-}STOCH.$$

In particular, any recognizable forest is ($\mathbb{K}$-) stochastic.

Proof. Let $G: T_\Sigma \to \mathbb{K}$ be a costable function and

$$F = \{t \in T_\Sigma \mid G(t) \neq 0\}.$$

Without any loss, we may assume that G takes non-negative values; indeed by Proposition 5.1 [1], the function

$$G \odot G: T_\Sigma \to \mathbb{K}, \quad (G \odot G)(t) = G^2(t) \geqslant 0$$

is costable and has the same support as G. Hence,

$$F = \{t \in T_\Sigma \mid G(t) > 0\}.$$

Consequently, according to the previous corollary,

$$F \in \mathbb{K}\text{-STOCH}.$$

Our last assertion comes from the fact that any recognizable forest is the support of a certain costable function. $\square$

Proposition 5. *Let $F_1 \in \mathbb{K}$-SUPP and $F_2 \in \mathbb{K}$-STOCH; then for any symbol $\sigma \in \Sigma_2$ the forest*

$$\sigma(F_1, F_2) = \{\sigma(t_1, t_2) \mid t_i \in F_i,\ i = 1, 2\}$$

(called σ-top-catenation of F_1 and F_2 [3]) belongs to $\mathbb{K}$-STOCH.

Proof. We have

$$F_1 = \{t \in T_\Sigma \mid G_1(t) \neq 0\}, \qquad F_2 = \{t \in T_\Sigma \mid G_2(t) > 0\}$$

with $G_1 : T_\Sigma \to \mathbb{K}$ costable ($i = 1, 2$) and $G_1(t) \geqslant 0$ for all $t \in T_\Sigma$. But as it is shown in [1, Proposition 6.5], the function

$$\sigma(G_1, G_2) : T_\Sigma \to \mathbb{K}$$

defined by

$$\sigma(G_1, G_2)(t) = \begin{cases} G_1(t_1) \cdot G_2(t_2) & \text{whenever } t = \sigma(t_1, t_2), \\ 0 & \text{otherwise} \end{cases}$$

is costable an we have

$$\sigma(G_1, G_2)(t) > 0$$

iff

$$t = \sigma(t_1, t_2) \quad \text{and} \quad G_1(t_1) \cdot G_2(t_2) > 0$$

iff

$$t = \sigma(t_1, t_2) \quad \text{and} \quad G_1(t_1) > 0,\ G_2(t_2) > 0$$

i.e. iff

$$t = \sigma(t_1, t_2) \quad \text{and} \quad t_1 \in F_1,\ t_2 \in F_2.$$

In other words,

$$\sigma(F_1, F_2) = \{t \mid \sigma(G_1, G_2)(t) > 0\}$$

and this completes the proof. $\square$

Proposition 6. *The intersection of a support with a stochastic forest, is again a stochastic forest.*

Proof. We omit the proof. □

Proposition 7. *The union of a forest $F \in STOCH$ with a forest $R \in \mathbf{Q}$-$SUPP$, is a stochastic forest, too.*

Proof. Let

$$F = \{t \mid G(t) > 0\}, \qquad R = \{t \mid G'(t) > 0\}$$

with $G: T_\Sigma \to \mathbb{R}$, $G': T_\Sigma \to \mathbf{Q}$ costable and $G'(t) \geqslant 0$ for all $t \in T_\Sigma$.

As we have seen in Proposition 1, there exists a real number $M > 0$ such that

$$|G(t)| < M^{1+\mathrm{size}(t)} \quad \text{for all } t \in T_\Sigma.$$

Thus, the function $\bar{G}: T_\Sigma \to \mathbb{R}$ with

$$\bar{G}(t) = \frac{G(t)}{M^{1+\mathrm{size}(t)}}, \quad t \in T_\Sigma$$

is costable, $F = \{t \mid \bar{G}(t) > 0\}$ and

$$|\bar{G}(t)| < 1 \quad \text{for all } t \in T_\Sigma. \tag{1}$$

On the other hand, by virtue of Proposition 2, from the function G' we can construct a costable function $\bar{G}': T_\Sigma \to \mathbf{Z}$ with the property

$$\{t \mid \bar{G}'(t) \neq 0\} = \{t \mid G'(t) \neq 0\}$$

and $\bar{G}(t) \geqslant 0$ for all t. We have

$$(\bar{G} + \bar{G}')(t) > 0 \text{ iff } \bar{G}(t) + \bar{G}'(t) > 0$$

from which, if we take into account (1) and the fact $\bar{G}'(t) \in \mathbf{Z}$ for all t, we equivalently get

$$\bar{G}(t) > 0 \quad \text{or } \bar{G}'(t) > 0,$$

that is,

$$F \cup R = \{t \mid (\bar{G} + \bar{G}')(t) > 0\}$$

and the proof is completed. □

In [8, Theorem III, 4.5] a **Q**-stochastic language L is provided, whose square LL fails to be **Q**-stochastic.

At a first point of view, therefore, we have to expect a few things about catenation closure.

However, the frontier technique allows to get some interesting results in this direction.

First of all, we recall that the function "frontier" $\mathrm{fr}: T_\Sigma \to \Sigma_0^*$ is recursively defined by

– $\mathrm{fr}(c) = c$ $(c \in \Sigma_0)$,

– $\mathrm{fr}(\sigma(t_1, \ldots, t_n)) = \mathrm{fr}(t_1) \cdots \mathrm{fr}(t_n)$,

that is, to any tree corresponds the word obtained by catenating its leaves from the left to the right.

Next the following holds:

Proposition 8. *For any rational function $r: \Sigma_0^* \to \mathbb{K}$, the composition $r \circ \mathrm{fr}: T_\Sigma \to \mathbb{K}$*

$$(r \circ \mathrm{fr})(t) = r(\mathrm{fr}(t)), \quad t \in T_\Sigma$$

is costable.

Consequently, the inverse image $\mathrm{fr}^{-1}(L)$ of a $\mathbb{K}$-stochastic language $L \subseteq \Sigma_0^*$ is a $\mathbb{K}$-stochastic forest.

Proof. The first half of the proposition has already been proved in [2].

Next, by virtue of (word) Turakainen's theorem

$$L = \{w \in \Sigma_0^* \mid r(w) > 0\}$$

with $r: \Sigma_0^* \to \mathbb{K}$ rational. Then $r \circ \mathrm{fr}$ is costable and

$$\begin{aligned} \mathrm{fr}^{-1}(L) &= \{t \in T_\Sigma \mid \mathrm{fr}(t) \in L\} \\ &= \{t \in T_\Sigma \mid r(\mathrm{fr}(t)) > 0\} \\ &= \{t \in T_\Sigma \mid (r \circ \mathrm{fr}(t) > 0\}. \end{aligned}$$

Hence $\mathrm{fr}^{-1}(L)$ is, by Theorem 2, a $\mathbb{K}$-stochastic forest. □

Let us denote by $\mathbb{K}$-SCf the family of languages that are frontiers of $\mathbb{K}$-stochastic forests (called $\mathbb{K}$-stochastic-context-free).

Combining the previous proposition with the obvious equality $L = \mathrm{fr}(\mathrm{fr}^{-1}(L))$, we immediately deduce the inclusion

$$\mathbb{K}\text{-Stoch} \subseteq \mathbb{K}\text{-SCf}. \tag{e}$$

Moreover

Proposition 9. *The catenation of a $\mathbb{K}$-stochastic-context-free language with a $\mathbb{K}$-rational language (i.e. a support of a $\mathbb{K}$-stochastic function [8]) is again a $\mathbb{K}$-stochastic-context-free language. Because of (e), this holds in particular for $\mathbb{K}$-stochastic languages.*

Proof. Let $L \subseteq \Sigma_0^*$ belong to $\mathbb{K}$-SCf and $M \subseteq \Sigma_0^*$ be $\mathbb{K}$-rational; then

$$L = \mathrm{fr}(F), \quad F \in \mathbb{K}\text{-STOCH} \text{ and } \mathrm{fr}^{-1}(L) \in \mathbb{K}\text{-SUPP}$$

so that, for any symbol $\sigma \in \Sigma_2$, the forest $\sigma(F, \mathrm{fr}^{-1}(L))$ belongs to $\mathbb{K}$-STOCH and therefore

$$ML = \mathrm{fr}[\sigma(F, \mathrm{fr}^{-1}(L))] \in \mathbb{K}\text{-SCf}. \quad \square$$

The same technique is applied to show

Proposition 10. (i) *The intersection of a* $\mathbb{K}$*-stochastic-context-free language with a* $\mathbb{K}$*-rational one, is again* $\mathbb{K}$*-stochastic-context-free.*

(ii) *The union of a* $\mathbb{K}$*-stochastic-context-free language with a* **Q***-rational one, is a stochastic-context-free language.*

Remark. Since by [5], the frontiers of the recognizable forests are exactly the context-free languages, similar to the above results can be stated by replacing "$\mathbb{K}$-rational" by "context-free".

References

[1] J. Berstel and C. Reutenauer, Recognizable formal power series on trees, *Theoret. Comput. Sci.* (1982) 113–148.
[2] S. Bozapalidis, Representable treeseries, *Fundam. Inform.*, in press.
[3] J. Engelfriet, Tree automata and tree grammars, Lecture Notes DIAMI, FN-10 Aarbus (1975).
[4] M. Fliess, Propriétés booléenes des languages stochastiques, *Math. Systems Theory* (1973) 353–359.
[5] F. Gesceg and M. Steinby, *Tree Automata* (Academiai Kiado, Budapest, 1984)
[6] O. Louscou-Bozapalidou, Stochastic tree functions, *Int. J. Comput. Math.*, in press.
[7] M. Magidor and G. Moran, Probabilistic tree automata, *Israel J. Math.* **8** (1970) 340–346.
[8] A. Salomaa and M. Soittola, *Automala-Theoretic Aspects of Formal Power Series* (Springer, Berlin, 1978).
[9] P. Turakainen, Generalized automata and stochastic languages, *Proc. Amer. Math. Soc.* **21** (1969) 303–309.
[10] P. Turakainen, Word-functions of stochastic and pseudo-stochastic automata, *Ann. Acad. Sci. Fenn. Ser. AI Math.* **1** (1975) 27–37.